# Analytical Comparison of the Genomic Mitochondrial DNA of Eleven Coregonines

T.L. Hisanaga, R. Bajno, T.J. Carmichael, and J.D. Reist

Central and Arctic Region
Fisheries and Oceans Canada
Winnipeg, MB R3T 2N6

2003

**Canadian Technical Report of Fisheries and Aquatic Sciences 2486**

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

**Canadian Technical Report of Fisheries and Aquatic Sciences**

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of the Department of Fisheries and Oceans, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in *Aquatic Sciences and Fisheries Abstracts* and indexed in the Department's annual index to scientific and technical publications.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and the Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report 925.

Technical reports are produced regionally but are numbered nationally. Request for individual reports will be filled by the issuing establishment listed on the front cover and title page. Out-of-stock reports will be supplied for a fee by commercial agents.

**Rapport technique canadien des Sciences halieutiques et aquatiques**

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. Il n'y aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques du ministère des Pêches et des Océans, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications complètes. Le titre exact paraît au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la revue *Résumés des sciences aquatiques et halieutiques*, et ils sont classés dans l'index annual des publications scientifiques et techniques du Ministère.

Les numéros 1 à 456 de cette série ont été publiés à titre de rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de rapports techniques , de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés a titre de rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur don't le nom figure sur la couverture et la page du titre. Les rapports épuisés seront fournis contre rétribution par des agents commerciaux.

104-01404/

# Erratum

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. 2003. Analytical comparison of the genomic mitochondrial DNA of eleven Coregonines. Can. Tech. Rep. Fish. Aquat. Sci. 2486: vi + 39p.

Page vi: line 1 and 22 should read "2003"

Page 2: line 9 should read "Hisanaga *et al.*, 2003".

Page 5: line 37 change authors name to "Kaukoranta"

Page 6: line 35 should read "(Murray and Reist, 2003)"

Page 7: line 15 should read "DeCicco" and "G. Low"

Page 8: line 8 change authors name to "Kaukoranta"

Page 29: line 3 figure 1 should read " Appendix 2"

Page 36: figure 14, "C. zenithicus 1" should replace "C. zenithicus 21"

Canadian Technical Report of
Fisheries and Aquatic Sciences 2486

2003

# ANALYTICAL COMPARISON OF THE GENOMIC MITOCHONDRIAL DNA OF ELEVEN COREGONINES

by

T.L. Hisanaga, R. Banjo, T.J. Carmichael, and J.D. Reist

Central and Arctic Region
Fisheries and Oceans Canada
Freshwater Institute
501 University Crescent
Winnipeg, Manitoba R3T 2N6

© Her Majesty the Queen in Right of Canada, 2003.
Cat. No. Fs 97-6/2486E ISSN 0706-6457

Correct citation for this publication:

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. 2003. Analytical comparison of the genomic mitochondrial DNA of eleven Coregonines. Can. Tech. Rep. Fish. Aquat. Sci. 2486: vi + 39p.

# TABLE OF CONTENTS

Page

ABSTRACT/RÉSUMÉ ..... vi

INTRODUCTION ..... 1

MATERIALS AND METHODS ..... 2

RESULTS AND DISCUSSION ..... 3

ACKNOWLEDGEMENTS ..... 7

REFERENCES ..... 7

## LIST OF TABLES

Table Page

1. Character matrix of variable nucleotides based on aligned 12S rRNA sequences. ... 10

2. Character matrix of variable nucleotides based on aligned 16S rRNA sequences. ... 11

3. Character matrix of variable nucleotides based on aligned ND1 sequences. ..... 12

4. Character matrix of variable nucleotides based on aligned ND2 sequences. ..... 13

5. Character matrix of variable nucleotides based on aligned COI sequences. ..... 15

6. Character matrix of variable nucleotides based on aligned COII sequences. ..... 16

7. Character matrix of variable nucleotides based on aligned ATPase6 sequences. ..... 17

8. Character matrix of variable nucleotides based on aligned COIII sequences. ..... 18

9. Character matrix of variable nucleotides based on aligned ND3 sequences. ..... 19

10. Character matrix of variable nucleotides based on aligned ND4L sequences. ..... 20

TABLE OF CONTENTS

11. Character matrix of variable nucleotides based on aligned ND4 sequences. .....21

12. Character matrix of variable nucleotides based on aligned ND5 sequences. .....23

13. Character matrix of variable nucleotides based on aligned ND6 sequences. .....25

14. Character matrix of variable nucleotides based on aligned cyt b sequences. .....26

15. Character matrix of variable nucleotides based on aligned D-loop sequences. .....28

LIST OF FIGURES

Figure Page

1. Gene organization and sequencing strategy for the Coregonine mitochondrial genomes. .....29

2. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the 12S rRNA mtDNA gene relationships among the ciscoes studied. .....30

3. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the 16S rRNA mtDNA gene relationships among the ciscoes studied. .....30

4. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND1 mtDNA gene relationships among the ciscoes studied. .....31

5. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND2 mtDNA gene relationships among the ciscoes studied. .....31

6. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COI mtDNA gene relationships among the ciscoes studied. .....32

7. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COII mtDNA gene relationships among the ciscoes studied. .....32

TABLE OF CONTENTS

8. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ATPase6 mtDNA gene relationships among the ciscoes studied. .....33

9. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COIII mtDNA gene relationships among the ciscoes studied. .....33

10. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND3 mtDNA gene relationships among the ciscoes studied. .....34

11. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND4L mtDNA gene relationships among the ciscoes studied. .....34

12. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND4 mtDNA gene relationships among the ciscoes studied. .....35

13. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND5 mtDNA gene relationships among the ciscoes studied. .....35

14. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND6 mtDNA gene relationships among the ciscoes studied. .....36

15. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the cyt *b* mtDNA gene relationships among the ciscoes studied. .....36

16. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the D-loop mtDNA gene relationships among the ciscoes studied. .....37

LIST OF APPENDICES

Appendix Page

1. Table detailing fish species and catch locations used in this study .....38

2. PCR and sequencing primers used in the analysis of Coregonine mitochondrial genomes. .....39

## ABSTRACT

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. 2002. Analytical comparison of the genomic mitochondrial DNA of eleven Coregonines. Can. Tech. Rep. Fish. Aquat. Sci. 2486: vi + 39p.

Taxanomic relationships among the Coregonines are presently confusing because they are based upon phenotypically plastic characters. This confusion is readily exemplified in the case of *Coregonus zenithicus* (shortjaw cisco), which is often mistaken for *Coregonus artedi* (lake cisco). In an attempt to identify genetic differences between *C. zenithicus* and *C. artedi*, the mitochondrial DNA of four *C. zenithicus* samples from two locations, and three *C. artedi* samples were sequenced in their entirety. The same was done for an additional four species of ciscoes (one individual each): *C. autumnalis* (baicalensis), *C. autumnalis* (pollan), *C. hoyi*, and *C. sardinella*. The gene-coding regions of these fish, along with the control region (D-loop), were analysed for genetic differences and their phylogenetic relatedness. While no differences were seen to occur between *C. artedi* and *C. zenithicus*, minor genetic variation occurred among the remaining new-world ciscoes (*C. autumnalis* (pollan), *C. hoyi*, *C. zenithicus*, and *C. artedi*) when compared to one another. A greater degree of variation was observed between the old-world (*C. autumnalis* (baicalensis) and *C. sardinella*) and the new-world ciscoes (*C. autumnalis* (pollan), *C. hoyi*, *C. zenithicus*, and *C. artedi*).

Key Words: Coregonus zenithicus, Coregonus artedi, taxonomic relationships, phylogenetic relationships.

## RÉSUMÉ

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. 2002. Analytical comparison of the genomic mitochondrial DNA of eleven Coregonines. Can. Tech. Rep. Fish. Aquat. Sci. 2486: vi + 39p.

Les relations taxinomiques entre les Corégonidés prêtent à confusion en ce moment parce qu'elles reposent sur des caractéristiques plastiques sur le plan phénotypique. *Coregonus zenithicus* (cisco à mâchoires égales), qui est souvent pris pour *Coregonus artedi* (cisco de lac), est un très bon exemple de cette confusion. Pour essayer de trouver des différences génétiques entre ces deux espèces, l'ADN mitochondrial de quatre échantillons de *C. zenithicus*, provenant de deux endroits, et de trois échantillons de *C. artedi* a été entièrement séquencé. Le séquençage a aussi été fait pour quatre autres espèces de ciscos (un spécimen de chaque espèce) : *C. autumnalis* (*baicalensis*), *C. autumnalis* (*pollan*), *C. hoyi* et *C. sardinella*. La séquence codante du gène chez ces poissons, ainsi que la boucle D, ont été analysées afin d'établir la présence de différences génétiques et le degré de relation phylogénétique. Bien qu'aucune différence n'ait été relevée entre *C. artedi* et *C. zenithicus*, une légère variation génétique a été identifiée entre les autres ciscos du Nouveau Monde (*C. autumnalis* (*pollan*), *C. hoyi*, *C. zenithicus* et *C. artedi*). Les ciscos de l'Europe (*C. autumnalis* (*baicalensis*) et *C. sardinella*) et ceux du Nouveau Monde (*C. autumnalis* [*pollan]*), *C. hoyi*, *C. zenithicus* et *C. artedi*) montraient un plus grand degré de variation.

Keywords: Coregonus zenithicus, Coregonus artedi, taxonomic relationships, phylogenetic relationships

# INTRODUCTION

The subfamily Coregoninae has proven difficult to resolve due to a high level of morphological plasticity and character convergence, especially among the North American (new-world) species (Todd and Smith, 1980). Furthermore, the current new-world ciscoes, including *Coregonus zenithicus* (shortjaw cisco) and *Coregonus artedi* (lake cisco), are geologically young, having only recently established themselves in their present areas following the retreat of the Pleistocene glaciers. It is likely the populations have not had sufficient time to establish genetic or population equilibrium (Elo, 1996), and that intra- and interspecific gene flow may be ongoing (Reist *et al.*, 1992). This suggests that differentiation will be relatively low, which in turn, suggests that characters which clearly delineate taxa may be lacking.

Based upon morphometric and meristic data, *C. artedi* and *C. zenithicus* have been classified as separate species, with gillraker number serving as the primary characters to identify species (Murray and Reist, 2003). *C. artedi* are associated with a higher gillraker number (35-48) than *C. zenithicus* (21-37). Gillraker number is one of the most important taxanomic characters among the Coregoninae and it has been considered to be genetically determined and not environmentally modifiable (Todd, 1998).

Although morphological studies have generally been successful in defining taxa, early fish taxonomy based upon morphometric and meristic data has led to the erroneous delineation of distinct species in some instances. For example, *C. nigripinnis cyanopterus* and *C. reighardi dymondi* were reclassified as being variants of *C. zenithicus* differentiated by spawning time and geographic location (Todd and Smith, 1980).

Mitochondrial DNA (mtDNA) has been widely used in phylogenetic studies to determine genetic relatedness both within taxa as well as among populations because it provides an abundance of genotypic characters analyzable through restriction fragment length analysis or polymerase chain reaction and subsequent nucleotide sequencing (Reed *et al.*, 1998; Reist *et al.*, 1998; Elo, 1996; Todd, 1981; and Avise, 1975). Other advantages to mtDNA analysis include its higher rate of evolution in comparison to nuclear DNA, and it is strictly maternally inherited with no homologous recombination (Moritz *et al.*, 1987; Avise and Vrijenhoek, 1987, Brown, 1983). Therefore, where current morphologically based taxonomy remains uncertain, molecular phylogenetics may be useful.

Although molecular phylogenetics has become a powerful tool in assessing genetic relatedness of species, it is not without limitations. Morphological systematics often employ characteristics that are readily visible such as the gillraker count as previously mentioned (Murray and Reist, 2003). The corresponding molecular counterpart for the differentiating characteristic may not be as readily apparent among the DNA sequences (for example a gene that controls gillraker count), and many sequences must be examined until an appropriate sequence discovered (Hillis, 1987). Such sequences may be found only in the nuclear genome where they can undergo homologous recombination (Moritz *et al.*, 1987), further increasing the difficulty in analyzing the changes. Sequence variation among individual nucleotides only show the change at a particular nucleotide at that point in time. A nucleotide may have undergone multiple changes at a given position, going from an A to a T, then either a C, or back to an A. Yet the change is often noted as only having occurred once, or in the case of returning to the

original nucleotide, as never having occurred in the first place. This is the primary limitation of numerical simulations. They always include gross simplifications of biological processes and the shortest steps to achieving the change is the one always assumed to have occurred (Hillis *et al.*, 1994).

Another problem in analyzing molecular data is the choice of method used to determine the molecular phylogeny. Substitution biases are complex, and the appropriate one-parameter (all substitutions are assumed to be equally probable), or two-parameter model (separate probabilities of substitution are assumed for transitions and transversions), must be chosen (Kimura, 1980). Although there exist many different analytical methods (Jukes-Cantor, Nei, Kimura, etc.), most programs are quite complex and require a good working knowledge of both statistics and substitution biases of the sequences in order to avoid choosing the wrong method. The method of tree construction is also important, as some utilize rate-dependent clustering algorithims (the unpaired group method with arithmetic mean), and others use a rate-dependent rooting algorithim, which may result in different trees (Hillis, 1987). Whatever the shortfalls of molecular or morphological systematics may be, the phylogenetic accuracy of a given species is strengthened when both methods are in agreement.

This report presents the analysis of mtDNA sequence data using a two-parameter method (Kimura, 1980) and presented as UPGMA (unweighted pair group method with arithmetic mean) phylograms (Felsenstein, 1993). The use of these molecular genetic techniques will hopefully provide additional information that will clarify the degree of distinctness and the taxanomic classification of Coregonine fish species, especially those that exhibit some level of risk.

## MATERIALS AND METHODS

Eleven coregonid fish (Appendix 1) had total DNA (mitochondrial and genomic) isolated from their tissue[1] using DNeasy tissue kits (Qiagen, Mississauga, ON.) as per manufacturer's recommendations and stored at –20°C. Polymerase chain reaction (PCR) mixtures were prepared containing 0.8 mM forward and reverse primers (Hisanaga *et al.*, in press), 0.1 mM dNTP, 1 X *Taq* polymerase buffer with 1.5 mM $MgCl_2$ (Roche Diagnostics, Laval, QC.), 10-60 ng of DNA, and 0.5 U *Taq* polymerase in a 100 µl reaction volume. Amplification of mtDNA was performed in a Techne Genius thermal cycler (Fisher Scientific, Ottawa, ON.) as per the thermal cycling parameters in Hisanaga *et al.*, in press. Successful amplification of the various mtDNA regions was verified by electrophoresis of 5 µl of the PCR product in a 1.0% agarose gel which was then stained with ethidium bromide (1 mg/ml).

Double stranded DNA products from PCR were purified by filtration through QIAquick columns (Qiagen, Mississauga, ON.) as per manufacturer's recommendations and stored at -20 °C until future use. The PCR products were then used for direct cycle sequencing with the dRhodamine dye-terminators (Applied Biosystems Inc, Foster City, CA) according to the manufacturer's instructions. Sequencing PCR parameters were as follows: initial denaturation at 94 °C for 4 minutes, and 25 cycles of: denaturation for 1

[1] Tissues can be muscle, fin clips, adipose fin, or any other tissue removed from either live-released of dead-sampled fish and preserved frozen in ethanol, or in a NaCl-saturated 20% dimethyl sulphoxide solution at 1:2 (v/v)

minute at 94 °C, annealing for 35 seconds at 53 °C, extension for 4 minutes at 72 °C. The primers used for target mtDNA amplification were also employed for sequencing. Additional primers (Invitrogen Canada, Burlington, ON.) were designed based upon previously sequenced segments in a process similar to chromosome walking. PCR and sequencing primer details, including location on the mtDNA genome and sequence are listed in Figure 1 and Appendix 2. Sequencing reactions were analyzed on an ABI Prism Model 377 DNA sequencer (Applied Biosystems, Foster City, CA).

Sequences were aligned and edited in MacVector 7.0 (Accelrys Inc., Madison, WI.). Further alignment of the gene-coding and control regions were performed using ClustalW (Thompson *et al.*, 1994). Phlyogenetic analyses were performed using PHYLIP software (Felsenstein, 1993). Distance matrices were calculated using the Kimura two-parameter model (Kimura, 1980), and distance trees were constructed following the UPGMA (unweighted pair group method with arithmetic mean. Autapomorphic insertions and deletions were ignored in analyzing the data, and all data were bootstrapped 100 times to ensure reliability of branch points.

## RESULTS AND DISCUSSION

Since total genomic DNA was amplified, caution was taken to ensure that the resulting sequences were not nuclear in origin, as insertion of mtDNA into the nuclear genome has been documented (Mindell *et* al., 1999). All DNA sequence electropherograms were examined for distinguishing features of nuclear copies such as double peaks resulting from potential co-amplification of mtDNA and nuclear DNA sequences, frameshifts, or stop codons. Features consistent with a mitochondrial origin include the presence of a conserved reading frame in protein coding genes and more importantly, the absence of extra stop codons, frameshifts, or unusual amino acid substitutions (Mindell *et al.*, 1999). The mtDNA genomic sequences were also quite similar to those of the only known published sequence of a Coregonine mitochondrial genome, *Coregonus lavaretus* (Miya and Nishida, 2000).

Genomic mtDNA sequences for the 11 fish were not completely obtained; some regions in individual fish were not resolved despite multiple attempts. For example, the ND1B primer was unable to provide a sequence for *C. zenithicus* 2, although it readily provided sequences for all the other fish. Due to the extremely high amount of similarity exhibited among *C. zenithicus*, *C. artedi*, and *C. hoyi*, it is unlikely that the primer was unable to bind to the template DNA. Instead, the quality of the template DNA for that particular fish was likely poor. There were some instances where individual base composition of the sequence could not be resolved, however, they are likely to be similar to the consensus sequence in the case of *C. artedi*, *C. zenithicus*, *C. hoyi*, and *C. autumnalis* (pollan) due to the strong degree of similarity among those individuals.

Large amounts of insertion-deletion (indels) and nucleotide changes were observed within the D-loop region sequence found in *C. zenithicus*, *C. hoyi*, *C. artedi*, *C. sardinella*, and *C. autumnalis* (baicalensis). The D-loop was aligned using ClustalW which yielded a minimum number of evolutionary steps with gaps inserted to maintain alignment. This was not visually apparent in the complete sequence alignments of all base pairs analyzed (data not shown).

Haplotype results for the various gene-coding and control regions for the eleven fish are given in Tables 1-15. The 12S rRNA region contains only three haplotypes, found in a group consisting of *C. sardinella*, *C. autumnalis* (baicalensis) and a group consisting of *C. artedi, C. hoyi, C. zenithicus* and *C. autumnalis* (pollan) forming the final haplotype (Table 1). The phylogenetic tree (Fig. 2) of the 12S rRNA indicates very little evolution has occurred, and that *C. autumnalis* (baicalensis) is more closely related to the new-world ciscoes (*C. artedi, C. hoyi, C. zenithicus*) than is *C. sardinella*. The results of the 16S rRNA region are depicted in Fig. 3 and Table 2. They show a single base pair deletion at position 1671 in *C. sardinella*, *C. autumnalis* (baicalensis) and *C. autumnalis* (pollan). The phylogenetic tree (Fig. 3) is divided into two clades, the first containing *C. sardinella* and *C. autumnalis* (baicalensis) and the second containing the remaining ciscoes, *C. artedi, C. hoyi, C. zenithicus* and *C. autumnalis* (pollan).

More variation is seen in the gene-coding region ND1 than in the rRNA regions (Table 3 and Fig. 4). Fish *C. zenithicus* 1 shares some of its base changes in common with *C. hoyi*, more so than with any of the other *C. zenithicus* and *C. artedi* to which it is more closely related. This is reflected in the phylogenetic tree (Fig. 4), in which two dominant branches are again seen, the first containing *C. sardinella* and *C. autumnalis* (baicalensis) and the second containing the remaining ciscoes, *C. artedi, C. hoyi, C. zenithicus* and *C. autumnalis* (pollan). The ND2 gene was not complete for *C. autumnalis* (baicalensis) and was omitted from analysis in Table 4 and Figure 5. The data remain consistent with that previously seen, in that *C. hoyi* and *C. zenithicus* 1 are very similar, and *C. sardinella* forms a distant branch from the other ciscoes. This gene demonstrates a high degree of variation for *C. autumnalis* (pollan) from the rest of the new-world ciscoes. Results for the COI gene (Table 6 and Fig. 6) follow those seen previously for the ND2 gene.

The COII gene sequence reveals very little difference among the new-world ciscoes. One exception is *C. zenithicus* 1 which contains an additional base change not seen between the new-world ciscoes and either *C. hoyi* or *C. autumnalis* (pollan) at position 315 in the sequence (Table 6 and Fig. 7). *C. sardinella* and *C. autumnalis* (baicalensis) form distinct branches from the remaining ciscoes with the branch for *C. sardinella* appearing as the most distant.

The ATPase6 gene exhibits the greatest amount of variation along the mitochondrial genome. This is readily seen in the number of haplotypes obtained (Table 7), eight out of the eleven fish are non-identical for this gene. The phylogram for ATPase6 is however similar to those shown previously, indicating no major genetic distinction among the new-world ciscoes.

The COIII, ND3, and ND4L genes are relatively short, and reveal little variation (Tables 8,9,10 and Figures 9, 10, 11), indicating the sequences are too short or too well conserved to be of any taxonomic value. Although quite a long gene, 1380 basepairs, the ND4 gene shows no variation among the *C. zenithicus* and *C. artedi*, except for *C. zenithicus* 1 which is most similar to *C. hoyi* (Fig. 12 and Table 11). Once again, *C. sardinella* and *C. autumnalis* (baicalensis) are distinct from the remainder, but in this instance they form a branching group distinct from the remaining ciscoes, and similar to the pattern observed for ND1 and 16S rRNA.

The longest gene-coding region in the mitochondrial genome, ND5, has proven informative in determining phylogenetic relatedness of Teleosteans (Miya and Nishida,

2000). However, no new information is derived for the ciscoes in this study (Table 12 and Fig. 13). While ND6 is slightly longer than COIII, ND3, and ND4L, it also provides insufficient variation to be considered taxonomically useful (Table 13, Fig. 14).Cytochrome *b*, despite its success in inferring relationships in other fish (Lydeard, and Roe, 1996), also provides no new information and appears to reveal very little variation among the new-world ciscoes analyzed in this study (Table 14 and Fig. 15).

The control region, or D-loop, is one of the most variable regions of the mitochondrial genome as it does not code for any genes, thereby eliminating the evolutionary constraints placed upon the other gene-coding regions. The D-loop sequences derived for the eleven fish studied are all shown containing a stretch of twelve thymidine nucleotides (Ts). This may not be a true representation of the number of Ts in that stretch, as the electropherograms indicate anywhere from 11-13. This result is probably due to the limitations of the sequencing procedure itself. *Taq* polymerase *in vitro* is without the proof reading capabilities and the mismatch repair mechanisms found in *in vivo* DNA replication (Saiki *et al*, 1988). During long stretches of repeating base sequence, slippage of the enzyme may occur, resulting in the addition or deletion of bases. This results in an incorrect representation of the number of times the base is repeated. Replication slippage also occurs *in vivo* and is an important source of naturally occurring mutation. However, due to the appearance of the electropherograms at this stretch, it was apparent that *in vitro* replication slippage resulted in the differing numbers of thymidine nucleotides. Due to the high degree of similarity to the other ciscoe species studied, to the sequences available from GenBank (Accession nos. U95183-95203), and to the *C. lavaretus* genome sequenced by Miya and Nishida (2000), a stretch of 12 Ts was concluded to exist in the eleven fish studied.

The complete D-loop sequences could not be determined for all eleven fish as sequences overlapping the tRNA phenylalanine gene were not obtained. Blast searches with the sequences that were obtained were in excellent agreement with those previously published (Reed *et al*., 1998). Phylogenetic analysis of the haplotypes (Table 15), resulted in a tree similar to those obtained from the aforementioned gene-coding regions (Fig. 16).

While the D-loop region was found to contain little information with regards to resolving the genetic distinctiveness between *C. artedi* and *C. zenithicus*, some surprising details were seen. *C. sardinella* and *C. autumnalis* (baicalensis) were very closely related, both sharing a large insertion block (Appendix 4). *C. autumnalis* (bacialensis) occurs only in Lake Baikal and the rivers feeding into it and has been shown to be genetically distinct from other species of *C. autumnalis* through electrophorectic analysis of enzymes (Bodaly *et al*., 1994) and by DNA hybridization (Kaukorante and Mednikov, 1988). This geographic isolation is likely responsible for the genetic differences to the North American corgonids. However, *C. sardinella* also contained the repeat blocks observed in *C. artedi, C. hoyi, C. zenithicus* and *C. autumnalis* (pollan). The D-loop and coding-region similarities shared between *C. sardinella* and *C. autumnalis* (baicalensis) indicated a close genetic and evolutionary relationship which is in good agreement with previous mitochondrial DNA restriction fragment analysis (Bernatchez and Dodson, 1991).

*C. autumnalis* (pollan) also contained many of the same conserved sequence blocks and repeats as found in *C. artedi, C. hoyi,* and *C. zenithicus*. This would indicate a common origin for these species. Comparison of a portion of the D-loop further

downstream of that previously investigated indicated that many of the coregonines designated as new-world (*C. artedi, C. zenithicus,* and *C. hoyi*) appeared to have an indel (i.e. insertion-deletion) of 17 bases (AAAACTTTTAATTATAC) occurring in the control region when compared to old-world coregonines (*C. sardinella,* and *C. autumnalis* (baicalensis). This sequence is similar to the LS promoter-like sequences (LSP, Saccone *et al.*, 1991). This indel also occured in the *C. autumnalis* (pollan) from Loch Neagh in Northern Ireland, suggesting ancestral forms of this coregonine may have been responsible for widespread re-population of the new-world during early glacial events. Although previous researchers have reported a close phylogenetic proximity between *C. autumnalis* (pollan) and *C. autumnalis* species from North America, (Bodaly *et al.*, 1994), this similarity between *C. autumnalis* (pollan) and the new-world ciscoes studied was unexpected and deserves further study.

All the individual phylograms based upon the various mtDNA regions sequenced generally agree with respect to the following points. First, both *C. sardinella* and *C. autumnalis* (baicalensis) were consistently more distant from the remaining taxa examined. Second, differences among *C. autumnalis* (pollan), *C. hoyi*, *C. artedi.* and *C. zenithicus* were generally small and variable, and relationships were not resolved. Third, two patterns of relationships of *C. sardinella* and *C. autumnalis* (baicalensis) to the new-world group were observed: one, in which these old-world taxa were each other's nearest relative (e.g. that portrayed by D-loop) as represented by nine phylograms; and one in which these two taxa were sequentially associated with the new-world group (e.g. that portrayed by 12S rRNA) as represented by four phylograms. In all of the latter cases, *C. sardinella* was the taxon most distant from all remaining taxa. A further two phylograms could not be summarized in this way since the *C. autumnalis* (baicalensis) data were missing due to a lack of PCR amplification.

In conclusion, despite exhaustive study, it would appear that the mitochondrial genome contains no phylogentically useful variation for the differentiation of *C. artedi* from *C. zenithicus*. This result agrees with previous studies of mtDNA using various approaches (Reed *et al.*, 1998, Sajdak and Phillips, 1997, Elo, 1996, Todd, 1981). Morphometric and meristic data therefore remain the best means of distinguishing *C. artedi* and *C. zenithicus* at the taxanomic level. However, studies of genomic DNA such as microsatellites may yet demonstrate consistent differences between these taxa.

As previously mentioned, the gillraker number is the key taxanomic character among the Coregoninae, and had previously been believed to be genetically determined and not environmentally modifiable (Murray and Reist, in press). However, laboratory-reared progeny of *C. artedi* had lower gillraker counts than their wild parents, whereas *C. zenithicus* laboratory-reared progeny had similar gillraker counts to the wild parents. The decrease in gillraker numbers was attributed to the lower rearing temperatures experienced by the fish in the laboratory than those in the wild (Todd, 1998). The depth range for *C. zenithicus* (20-180 m) is greater than that of *C. artedi* (up to 50 m) (Todd, 1998). As a result, *C. zentihicus* would experience a colder temperature than *C. artedi*, and this lake depth-temperature difference could be responsible for an ecologically plastic species to phenotypically appear as a species complex. Despite these results the most conservative present approach is to continue to recognise these taxa as specifically distinct until demonstrated otherwise.

With the recently undertaken sequencing project of the nuclear genome of Zebrafish (The Institute for Genomic Research), the potential for determining a rapidly evolving region of nuclear DNA that will be of use for phylogenetic analysis is fast approaching. A previous attempt to determine phylogenetic relatedness among the subfamily Coregoninae with a nuclear gene (Sajdak and Phillips, 1997) was unable to detect differences between *C. artedi* and *C. zenithicus*. Should the gillraker number indeed be genetically determined, and not environmentally modifiable, the gene(s) involved may be determined through the Zebrafish sequencing project. The differences in that/those gene(s) among *C. artedi* and *C. zentihicus* would be expected to be of taxonomic usefulness in genetically differentiating the two species.

## ACKNOWLEDGEMENTS

This study was supported by funding from the Department of Fisheries & Oceans Canada through its internal Species At Risk programme. The following individuals are thanked for providing or aiding in the collection of samples: K. Alt, D. Bodaly, F. Decicco, P. Gudkov, J. Johnson, A. Kaupinis, G. low, L. Murray, M. Papst, Y. Reshetnikov, M. Steinhilber, T. Todd, J. Vuorinen

## REFERENCES

Avise, J.C. 1975. Systematic value of electrophoretic data. Syst. Zool. 23: 465-481.

Avise, J.C., and R.C. Vrijenhoek. 1987. Mode of inheritance and variation of mitochondrial DNA in hybridogenetic fishes of the genus *Poeciliopsis*. Mol. Biol.Evol. 4: 514-525.

Bernatchez, L, F. Colombani, and J.J. Dodson. 1991. Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. J. Fish Biol. 39: 283-290.

Bodaly, R.A., D.A. Vuorinen, Y.S. Reshetnikov, and J.D. Reist. 1994. Genetic relationships of five species of coregonid fishes from Siberia. J. Icthyol. 34; 117-130.

Brown, W.M. 1983. Evolution of animal mitochondrial DNAs. *In* "Evolution of genes and proteins" (M. Nei and R.K. Koehn, Eds.). pp. 62-88. Sinauer, Sunderland, MA.

Elo, K. 1996. Intra- and interspecific gene flow and phylogenetic relationships in the Salmonidae as revealed by protein and DNA analyses. Ph. D. Thesis. University of Joensuu. Dept. Biology, Publications in Sciences No. 37, Joensuu, Finland.

Felsenstein, J. 1993. PHYLIP (phylogeny interface package) version 3.5c. *distributed by the author*. Department of Genetics, University of Washington, Seattle.

Hillis, D.M. 1987. Molecular versus morphological approaches to systematics. Ann. Rev. Ecol. Syst. 18: 23-42.

Hillis, D.M., J.P. Huelsenbeck, and C.W. Cunningham. 1994. Application and accuracy of molecular phylogenies. Science 264: 671-677.

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. In press. Design of primers for polymerase chain reaction amplification of mitochondrial DNA of Coregonine fishes. Can. Tech. Rep. Fish. Aquat. Sci. 2487.

Kaukorante, M., and B.M. Mednikov. 1988. Genetic divergence of whitefishes of Fenno-Scandia. *In* "Biologiya sigovykh rhb" (Biology of coregonids). pp. 31-48. Moscow, Nauka.

Kimura, M. 1980. A simple method for estimating evolutionary distances of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol.16: 111-120.

Lydeard, C. and K.J. Roe. 1996. The phylogenetic utility of the mitochondrial cytochrome *b* gene for inferring relationships among actinopterygian fishes. *In* "Molecular systematics of Fishes" (T.D. Kocher and C.A. Stepiens, Eds.). pp. 285-303. Academic Press, San Diego.

Mindell, D.P., M.D. Sorenson, D.E. Dimcheff, M. Hasegawa, J.C. Ast, and T. Yuri. 1999. Interordinal relationships of birds and other reptile based on whole mitochondrial genomes. Syst. Biol. 48: 138-152.

Miya M., and M. Nishida. 2000. Use of mitogenetic information in teleostean molecular phylogenetics: a tree based exploration under the Maximum-Parsimony optimality criterion. Mol. Phylogenetic Evol. 17: 437-455.

Murray, L., and J. D. Reist. 2003. Status report on the shortjaw cisco (*Coregonus zenithicus*) in central and western Canada. Can. Manuscr. Rep. Fish. Aquat. Sci. 2638 :vii + 56 p.

Moritz C., T. Dowling, and W.M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance of population biology and systematics. Ann. Rev. Ecol. Syst. 18: 269-292.

Reed, K.M., M.O. Dorschner, T.N. Todd, and R.B. Phillips. 1998. Sequence analysis of the mitochondrial DNA control region of ciscoes (genus *Coregonus*): taxonomic implications for the Great Lakes species flock. Mol. Ecol. 7: 1091-1096.

Reist, J.D., L.D. Maiers, R.A. Bodaly, J.A. Vuorinen, and T.J. Carmichael. 1998. The phylogeny of new- and old-world coregonine fishes as revealed by sequence

variation in a portion of the D-loop of mitochondrial DNA. Arch. Hydrobiol. Spec. Issues Advanc. Limnol. 50: 323-339.

Reist, J.D., Vuorinen, J., and Bodaly, R.A. 1992. Genetic and morphological identification of coregonid hybrid fishes from Arctic Canada. Pol. Arch. Hydrobiol. 39: 551-561.

Saccone, C., G. Pesole, and E. Sbisa. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. 33: 83-91.

Saiki, R.K., D.H. Gelfand, S. Stoffel, J. Scharf, R. Higuchi, G.T. horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Sajdak, S.L, and R.B. Phillips. 1997. Phylogenetic relationships among *Coregonus* species inferred from the DNA sequence of the first internal transcribed spacer (ITS1) of ribosomal DNA. Can. J. Fish. Aquat. Sci. 54: 1494-1503.

Thompson, J.D. D.G. Higgins, and T.J. Gibson. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nuc. Acids Res. 22: 4673-4680.

Todd, T.N. 1998. Environmental modification of gillraker number in coregonine fishes. Ergebnisse-der-Limnologie. 50: 305-315.

Todd, T.N. 1981. Allelic variabilities in species and stocks of Lake Superior ciscoes (*Coregoninae*). Can. J. Fish. Aquat. Sci. 38: 1808-1813.

Todd, T.N., and Smith G.R. 1980. Differentation of *Coregonus zenithicus* in LakeSuperior. Can. J. Fish. Aquat. Sci. 37: 2228-2235.

Table 1. Character matrix of variable nucleotides based on aligned 12S rRNA sequences. Numbers indicate position of sites in the 947 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | 123788<br>078867<br>779824 |
|---|---|
| Consensus | CAATCC |
| *C. artedi* 1 | ..... |
| *C. artedi* 2 | ..... |
| *C. artedi* 3 | ..... |
| *C. zenithicus* 1 | ..... |
| *C. zenithicus* 2 | ..... |
| *C. zenithicus* 3 | ..... |
| *C. zenithicus* 4 | ..... |
| *C. autumnalis* (baicalensis) | T·CC·· |
| *C. sardinella* | ·GCCAT |
| *C. hoyi* | ..... |
| *C. autumnalis* (pollan) | ..... |

Table 2. Character matrix of variable nucleotides based on aligned 16S rRNA sequences. Numbers indicate position of sites in the 1679 base pair alignment. Sites identical to the consensus sequence are indicated as a period, and deletions/insertions are indicated as dashes.

| | |
|---|---|
| | 111111111 |
| | 11111155568889902233467 |
| | 11247802285692642554707 1 |
| | 69045985779898922017 1314 |
| Consensus | TAATGACCTGCCTGCTCGTGGGGC |
| *C. artedi* 1 | ..... |
| *C. artedi* 2 | ..... |
| *C. artedi* 3 | ..... |
| *C. zenithicus* 1 | ..... |
| *C. zenithicus* 2 | ..... |
| *C. zenithicus* 3 | .....A.... |
| *C. zenithicus* 4 | ..... |
| *C. autumnalis* (baicalensis) | CCCCA.T..AATCAT.GAC.A.-- |
| *C. sardinella* | CCCC.GTAC...CA.CG.C...-T |
| *C. hoyi* | .....G..... |
| *C. autumnalis* (pollan) | .....A-. |

Table 3. Character matrix of variable nucleotides based on aligned ND1 sequences. Numbers indicate position of sites in the 1020 base pair alignment. Sites identical to the consensus sequence are indicated as a period. *C. zenithicus* 2 was omitted due to an incomplete sequence.

| | |
|---|---|
| | 1112222223333333334444445555556666667777778888888999 |
| | 1688936703477911124457802578900135602347934456813368890447 |
| | 5973473824713137238402121839795473753640823517359447428765 85 |
| Consensus | TGTGCGGCTGCCACGCAGAGTAGGGCCAGACTGTGGCCAGAGATGAATGTAATCGTGCAA |
| *C. artedi* 1 | ····A····· |
| *C. artedi* 2 | ·····A····· |
| *C. artedi* 3 | ·····A····· |
| *C. zenithicus* 1 | CA·····G· |
| *C. zenithicus* 3 | ····· |
| *C. zenithicus* 4 | ····· |
| C. *autumnalis* (baicalensis) | TGCA·A···AAT·TTTG····TTC···C·GT·ACAATGGAGAGCAC·CACGGCTACAT·G |
| *C. sardinella* | ·····ATTCAG··TT···GAC·TCATTCCGTC·CA·TGGAGAGCACACACGGCTACAT·G |
| *C. hoyi* | C·····G·····G· |
| *C. autumnalis* (pollan) | ·····G·····A·····T·····G····· |

Table 4. Character matrix of variable nucleotides based on aligned ND2 sequences. Numbers indicate position of sites in the 1050 base pair alignment. Sites identical to the consensus sequence are indicated as a period. Note that *C. autumnalis* (baicalensis) was omitted due to an incomplete sequence.

| | |
|---|---|
| | 1111122222333333444444444445555555555555666666677777777777778888 |
| | 6790228913449344588023356788990033346777779015678812223445566781113 |
| | 4568699237097028112862305736281415764034698054935750362176925690131 |
| Consensus | ACGTCTGTGAGGAAGTGATAAGATTTGGCTACGCAACGTACCATTACAAAGCCGACAGGGTAAAGA |
| *C. artedi* 1 | ···· |
| *C. artedi* 2 | ···· |
| *C. artedi* 3 | ···· |
| *C. zenithicus* 1 | ·····A·····A· |
| *C. zenithicus* 2 | ···· |
| *C. zenithicus* 3 | ···· |
| *C. zenithicus* 4 | ···· |
| *C. sardinella* | GTCATCACAAAGGGAAAGCTAAGCCCATTCAAGTAAGGTACCTGGCCGAGGCATTAGGCAAGGGAG |
| *C. hoyi* | ·····A·····A· |
| *C. autumnalis* (pollan) | ····A··G·A··A·A···G·····A·····A·····A· |

Table 4. Continued

| | |
|---|---|
| | 111111 |
| | 88888899999999999000000 |
| | 55578813456889990134 45 |
| | 23506886216120342754 70 |
| Consensus | TAGCGTATCATCCCGTGTGCGG |
| *C. artedi* 1 | ..... |
| *C. artedi* 2 | ..... |
| *C. artedi* 3 | ..... |
| *C. zenithicus* 1 | ·G····· |
| *C. zenithicus* 2 | ..... |
| *C. zenithicus* 3 | ..... |
| *C. zenithicus* 4 | ..... |
| *C. sardinella* | A·ATACCCTGCGTTACACCTAA |
| *C. hoyi* | ·····C····· |
| *C. autumnalis* (pollan) | ·G····· |

Table 5. Character matrix of variable nucleotides based on aligned COI sequences. Numbers indicate position of sites in the 1552 base pair alignment. Sites identical to the consensus sequence are indicated as a period. Note that *C. autumnalis* (baicalensis) was omitted due to an incomplete sequence.

| | |
|---|---|
| | 11111111111111111111111111111 |
| | 11233345566666667778888999999900000011112222223333333344445 |
| | 35528982945789991560178000344893668882359004899023344791484 |
| | 20017005139540697927608069058109253798427092409281417175068 2 |
| Consensus | TTTGCTATTTTGGGATCATTTTTTTCATAATCGACTACCGATTCTGCTATCCTATATTTA |
| *C. artedi* 1 | . . . . . |
| *C. artedi* 2 | . . . . . |
| *C. artedi* 3 | . . . . . |
| *C. zenithicus* 1 | . . . . . T . . . . . A . . . . . C . . . . . C . . . . . |
| *C. zenithicus* 2 | . . . . . |
| *C. zenithicus* 3 | . . . . . |
| *C. zenithicus* 4 | . . . . . |
| *C. sardinella* | CCCC . CGCCCCAA . GCTGCGACGCCAG . GGCTACTCGT . AGC . TC . TCGCTTCGCGCCCG |
| *C. hoyi* | . . . . . T . . . . . C . . . . . C . . . . . |
| *C. autumnalis* (pollan) | . . . . . C . . . . . C . . . . . C . . . . G . . . C . . A . . . . . |

Table 6. Character matrix of variable nucleotides based on aligned COII sequences. Numbers indicate position of sites in the 691 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | |
|---|---|
| | 1122222223333334444444555 |
| | 6611367890013890012568178 |
| | 2836413510959502619923335 |
| Consensus | CCCCCATCTAGCTGCCCGACTCAAA |
| *C. artedi* 1 | ····· |
| *C. artedi* 2 | ····· |
| *C. artedi* 3 | ····· |
| *C. zenithicus* 1 | ··T·····T····· |
| *C. zenithicus* 2 | ····· |
| *C. zenithicus* 3 | ····· |
| *C. zenithicus* 4 | ····· |
| *C. autumnalis* (baicalensis) | T·TT··CTCG··C·TGTAGACT·GG |
| *C. sardinella* | TTTTAGCTCAA·CA··TAG···GG· |
| *C. hoyi* | ··T····· |
| *C. autumnalis* (pollan) | ··T····· |

Table 7. Character matrix of variable nucleotides based on aligned ATPase6 sequences. Numbers indicate position of sites in the 684 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | |
|---|---|
| | 111111111122222222223333333333344444445555 66 |
| | 455881445556788011445579902233455678001244915884 5 |
| | 814141470392736469395634794709814081581947835382 7 |
| Consensus | TTAAAAAACCGCCCAGAATAGTTGCAAAAGCTGCGGACACTTAAGCCGG |
| *C. artedi* 1 | ·····C····· |
| *C. artedi* 2 | ····· |
| *C. artedi* 3 | ····G····· |
| *C. zenithicus* 1 | ·····A·····T····· |
| *C. zenithicus* 2 | ·····G····· |
| *C. zenithicus* 3 | ····· |
| *C. zenithicus* 4 | ····· |
| *C. autumnalis* (baicalensis) | CC·G·GGGT·A·T·G·GGCGACCA·G··GA··A··AT·C··CGG·A·CA |
| *C. sardinella* | ··C··G·GTTATTA·AGG·GA·CAAG·G·AT·A·AGCT·TCCG·AATGA |
| *C. hoyi* | ·····A····· |
| *C. autumnalis* (pollan) | ·····T·AT·····A····· |

Table 8. Character matrix of variable nucleotides based on aligned COIII sequences. Numbers indicate position of sites in the 786 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | |
|---|---|
| | 1122222223333333333344444455555555555666666666 |
| | 5391124556002223568992356890012335677701155788 |
| | 2553620251691470794069892351208472106965614810 |
| Consensus | TATTTCATGTTTCTACACGGATGATAACCAGTACGTACAATATTC |
| *C. artedi* 1 | ····A····· |
| *C. artedi* 2 | ····A····· |
| *C. artedi* 3 | ····A····· |
| *C. zenithicus* 1 | ····A····· |
| *C. zenithicus* 2 | ····A····· |
| *C. zenithicus* 3 | ····A····· |
| *C. zenithicus* 4 | ····A····· |
| *C. autumnalis* (baicalensis) | CGCCTTC·ACCC·CG·GTA·GCAGCG·TT·C·GGAC·TCCCCCCT |
| *C. sardinella* | C·CC··CGAC··GC·G·TA··CAG··GTTGCCGGACGTCCCC·C· |
| *C. hoyi* | ····A····· |
| *C. autumnalis* (pollan) | ····A····· |

Table 9. Character matrix of variable nucleotides based on aligned ND3 sequences. Numbers indicate position of sites in the 349 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | 1111112222223<br>5790133552255790<br>1395758362815343 |
|---|---|
| Consensus | ATGGGGACCACAAGAT |
| *C. artedi* 1 | ················ |
| *C. artedi* 2 | ················ |
| *C. artedi* 3 | ················ |
| *C. zenithicus* 1 | ··A············· |
| *C. zenithicus* 2 | ················ |
| *C. zenithicus* 3 | ················ |
| *C. zenithicus* 4 | ················ |
| *C. autumnalis* (baicalensis) | GC·A·A·TTG·CC·GC |
| *C. sardinella* | ·C··AAGTT·TCC·GC |
| *C. hoyi* | ··A············· |
| *C. pollan* | ················ |

Table 10. Character matrix of variable nucleotides based on aligned ND4L sequences. Numbers indicate position of sites in the 297 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | 111111112222<br>133788893458<br>702436721925 |
|---|---|
| Consensus | CCGGGGCAGGGT |
| *C. artedi* 1 | ............ |
| *C. artedi* 2 | ............ |
| *C. artedi* 3 | ............ |
| *C. zenithicus* 1 | ............ |
| *C. zenithicus* 2 | ............ |
| *C. zenithicus* 3 | ............ |
| *C. zenithicus* 4 | ............ |
| *C. autumnalis* (baicalensis) | T·AA·AT·AAAC |
| *C. sardinella* | TTAAAATGAAA· |
| *C. hoyi* | ............ |
| *C. autumnalis* (pollan) | ·····AT····· |

Table 11. Character matrix of variable nucleotides based on aligned ND4 sequences. Numbers indicate position of sites in the 1380 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | 1111122222444444455555555555666666667777777788888888888 |
|---|---|
| | 2345777945888678990256666012334677880334468901245669001225577 88 |
| | 7934235340469138478660235195479469286365894357940082146582869 29 |
| Consensus | AAACGGGTTATAGCCCCTGGTTTCGCGGAGATACGAACGCGGTCTGCAACGCTAATCAGGACC |
| *C. artedi* 1 | ····· |
| *C. artedi* 2 | ····· |
| *C. artedi* 3 | ····· |
| *C. zenithicus* 1 | ·····T···T·····A··· |
| *C. zenithicus* 2 | ····· |
| *C. zenithicus* 3 | ····· |
| *C. zenithicus* 4 | ····· |
| *C. autumnalis* (baicalensis) | ·CGTA·ACC·CGTAT·TC·A·CCTAA··GA··GTAGG·CTAAC·C·T·G·AACGGCTG·AGAT |
| *C. sardinella* | GC··A·ACCG·AA·TTTCAAC·CT·TA·GAGC·T····C···C·CA·GGT·····CT··CG·· |
| *C. hoyi* | ·····T···T·····AA··· |
| *C. autumnalis* (pollan) | ·····A·····T·····A·····T···T·····T·····A···· |

Table 11. Continued.

| | 1111111111111111111111111111 |
|---|---|
| | 8999900000111111111111222223333333 |
| | 9056656678011224566779025690002455 |
| | 4023963579138149814036344002689103 |
| Consensus | AGATTCCTAGTTGGGGCACTGCCGAGGTGGCCAG |
| *C. artedi* 1 | ···· |
| *C. artedi* 2 | ···· |
| *C. artedi* 3 | ···· |
| *C. zenithicus* 1 | ··G···· |
| *C. zenithicus* 2 | ···· |
| *C. zenithicus* 3 | ···· |
| *C. zenithicus* 4 | ···· |
| *C. autumnalis* (baicalensis) | C·GC·T·C·ACCAA····TCA·A·CAAC·AT··A |
| *C. sardinella* | CA·CCTTCG·ACAAAATGTCATGAT·ACA··T·G |
| *C. hoyi* | ··G·····G· |
| *C. autumnalis* (pollan) | ··A··T·····A·····C····· |

Table 12. Character matrix of variable nucleotides based on aligned ND5 sequences. Numbers indicate position of sites in the 1838 base pair alignment. Sites identical to the consensus sequence are indicated as a period, and unknown bases as X.

| | |
|---|---|
| | 111111 |
| | 1111112222333333444444444555555556666666777888899999999000000 |
| | 7001568268857899901223788922233457001236915727994556667881224 55 |
| | 6344080278975403651695409225847984092150979780475470392174691 03 |
| Consensus | TAGTAGTACCGGCGGACCGATGCTACGGCGTGAGCTGCAATAAATCTAGGAACTGGACATAGA |
| *C. artedi* 1 | ..... |
| *C. artedi* 2 | ..... |
| *C. artedi* 3 | ..... |
| *C. zenithicus* 1 | ..... |
| *C. zenithicus* 2 | ..... |
| *C. zenithicus* 3 | ..... |
| *C. zenithicus* 4 | ..... |
| *C. autumnalis* (baicalensis) | CGACGA·GCTAA··A··TAGC··C··AA··CAG·T·A····TGGCTCGAA·G·CA·G··C·A· |
| *C. sardinella* | C·ACGACGTTA·TAAGT··GCATCGTA·TACAG·TCAT·GCTCGC·CGA·GGTCAAGT·CGAT |
| *C. hoyi* | ..... |
| *C. autumnalis* (pollan) | ..... |

Table 12. Continued.

```
                          11111111111111111111111111111111111111111111111111111111111111111
                          00001111112222223333333344444444455555555556666666666777777777 8
                          77880144684557990155678901234577700135568991112255568133467799 1
                          47097936725475395759877540516316969231731371280703620617143924 9

Consensus                 TGAACCTCCATCCAAGTATCTACATGTCAGAGGACTAGAACTCCAATCCATCGGCAGGAGTAG

C. artedi 1               ·····························································
C. artedi 2               ·····························································
C. artedi 3               ·····························································
C. zenithicus 1           ·····························A·····························
C. zenithicus 2           ·····················GC··········G··························
C. zenithicus 3           ·····························································
C. zenithicus 4           ·····················X·····C·····X·····C·····X·····C·····X·····
C. autumnalis (baicalensis) ····························G····························
C. sardinella             ··GG···TT·ATAGGACG··CGTGCAC···GAA·T·GAGGTCT·G·C··CCTAATCAAGAC·A
C. hoyi                   C·G··TCT·GA·A·G·CG·T·G···A·TGAG··GT·GAG··CTTGG·T·G····T··AGACG·
C. autumnalis (pollan)    ··G·····························GC···························
                          ·A··T·····G·····A·····GG·····
```

Table 13. Character matrix of variable nucleotides based on aligned ND6 sequences. Numbers indicate position of sites in the 522 base pair alignment. Sites identical to the consensus sequence are indicated as a period.

| | 1111111122222222233333333444444444444455 |
|---|---|
| | 6789155677881234699992333568911222356778911 |
| | 6880737278367279756787039791917679802450205 |
| Consensus | TGACACCGCCGTTGTCGTCGGTGGACCGACGACAATATGTAAG |
| *C. artedi* 1 | ····· |
| *C. artedi* 2 | ····· |
| *C. artedi* 3 | ·····A····· |
| *C. zenithicus* 1 | ·····T····· |
| *C. zenithicus* 2 | ····· |
| *C. zenithicus* 3 | ····· |
| *C. zenithicus* 4 | ····· |
| *C. autumnalis* (baicalensis) | AAGTGTTAA··CA·CTACTC···AG·A·GAAC···C·C·CGGA |
| *C. sardinella* | ·A····T··TAC·AC·AC···CAAG·AACAACA·G·GCACGGA |
| *C. hoyi* | ·····T····· |
| *C. autumnalis* (pollan) | ·····G·····G· |

Table 14. Character matrix of variable nucleotides based on aligned cyt *b* sequences. Numbers indicate position of sites in the 1125 base pair alignment (the first 15 bases of the sequences are not included). Sites identical to the consensus sequence are indicated as a period.

| | |
|---|---|
| | 11122233333333444444445555566666666666777777777788888888899999 |
| | 55239235012578891223578905689022233458990235578899001245789022 3 |
| | 25687545984424707395919278751914703277032083646928473598327914 3 |
| Consensus | AGTTGTATCAGCACATCAACAAGTTTTTCATGGACTAAGGTGAATAACTACGGGATCGTCTTA |
| *C. artedi* 1 | ..... |
| *C. artedi* 2 | .....A..... |
| *C. artedi* 3 | ..... |
| *C. zenithicus* 1 | ..... |
| *C. zenithicus* 2 | ..... |
| *C. zenithicus* 3 | ..... |
| *C. zenithicus* 4 | ..... |
| *C. autumnalis* (baicalensis) | CTCCACT··C····GCT···GGACCC·C··CAAGT·GG·A·AGGCG··CC···C·C·AC·ACG |
| *C. sardinella* | C·CC·CCCTCAAG··CTGTTCGACCC·CGG·A·G·C··AACAA·C·GTCCTAAAGCACAT··G |
| *C. hoyi* | .....C..... |
| *C. autumnalis* (pollan) | C.....AA.....C..... |

Table 14. Continued.

| | 111111111<br>9999999000000011<br>3446789334467923<br>6233243284728221 |
|---|---|
| Consensus | AGTAGATGTTTCACCC |
| *C. artedi* 1 | ···· |
| *C. artedi* 2 | ···· |
| *C. artedi* 3 | ···· |
| *C. zenithicus* 1 | ···· |
| *C. zenithicus* 2 | ···· |
| *C. zenithicus* 3 | ···· |
| *C. zenithicus* 4 | ···· |
| *C. autumnalis* (baicalensis) | GACGA·CACCC··TT· |
| *C. sardinella* | G·CGAGCACCCTG·TT |
| *C. hoyi* | ···· |
| *C. autumnalis* (pollan) | ····A····· |

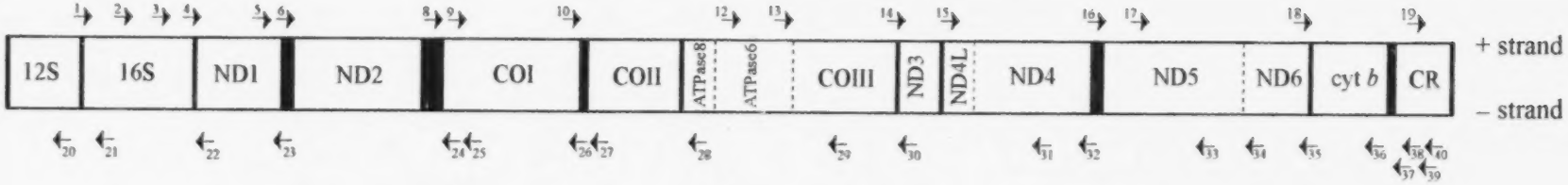
Table 15. Character matrix of variable nucleotides based on aligned D-loop sequences. Numbers indicate position of sites in the 1140 base pair alignment. Sites identical to the consensus sequence are indicated as a period, and deletions/insertions are indicated as dashes.

```
                                                                                                                  11111111111
                                  1111112222233344455666777777777888999999999999999999900000000000
                                  2390124682235856688936377011135889688000122346666789990222222455 5
                                  1940741207769066907786802577135640499281498592804892516732345695146

Consensus                         GATCTTGCATTCGAACAATCAGGTACATTGTCCGCGTCTGATATTAATATTATCAACACTTAATTCA

C. artedi 1                       ...............................................................................
C. artedi 2                       ...T...........................................................................
C. artedi 3                       ...............................................................................
C. zenithicus 1                   .......A..........G.............T..................G..............
C. zenithicus 2                   .....G.........................................................................
C. zenithicus 3                   ...............................................................................
C. zenithicus 4                   .......A.......................................................................
C. autumnalis (baicalensis)       AGCT··A·T-----·AGGA··TC·CTGCC·CT·AT·C····CGCC····C···ATGATTAAC·CCG·
C. sardinella                     A·CTC·A··C-AAGGA···ATT·C·TGCCT·T·ATAC·CAGCGCCGGCGCC·C·T·ATTAACGC··G
C. hoyi                           ...............................................G...............
C. autumnalis (pollan)            ..........................................T.............G..............
```

Figure 1. Gene organization and sequencing strategy for the Coregonine mitochondrial genomes. All protein-coding regions are encoded by the +strand with the exception of ND6, which is coded by the -strand. Transfer RNA genes are depicted as solid lines, with dashed lines depicting the boundaries of genes. Relative primer positions are indicated with numerals that correspond to Appendix 1. 12S and 16S rRNA are the 12S and 16S ribosomal RNAs; ND1-6 and ND4L are the NADH dehydrogenase subunits 1-6 and 4l; ATPase6 and 8 are the ATPase subunits 6 and 8; COI-III are the cytochrome oxidase subunits 1-3; cyt b is the cytochrome b; and CR is the control region.



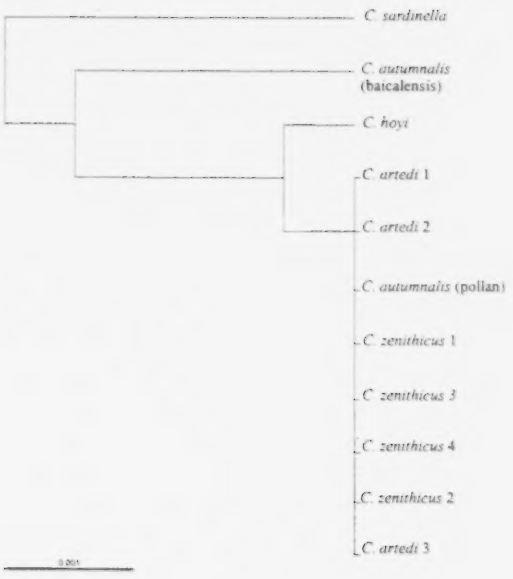



Figure 2. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the 12S rRNA mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

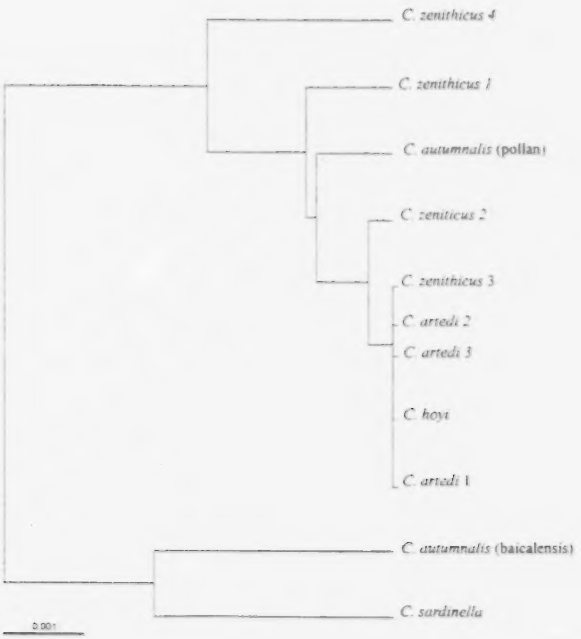



Figure 3. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the 16S rRNA mtDNA relationships among the ciscoes. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

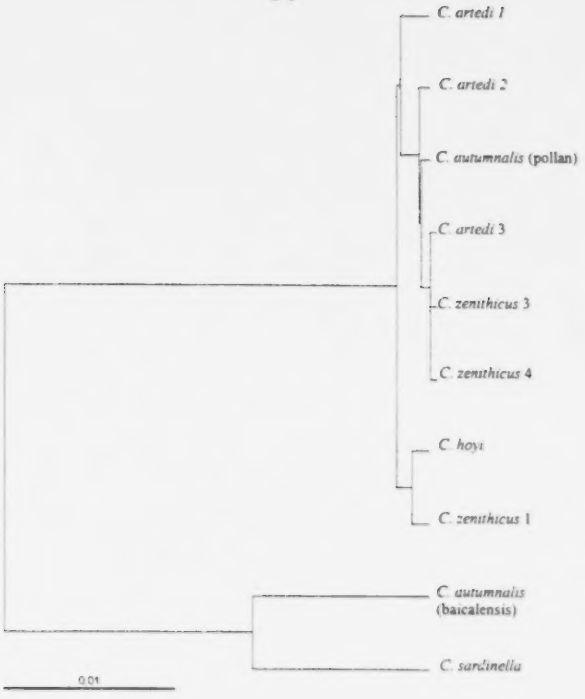



Figure 4. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND1 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter method (Kimura, 1980). Note that *C. zenithicus* 2 was omitted due to an incomplete sequence.

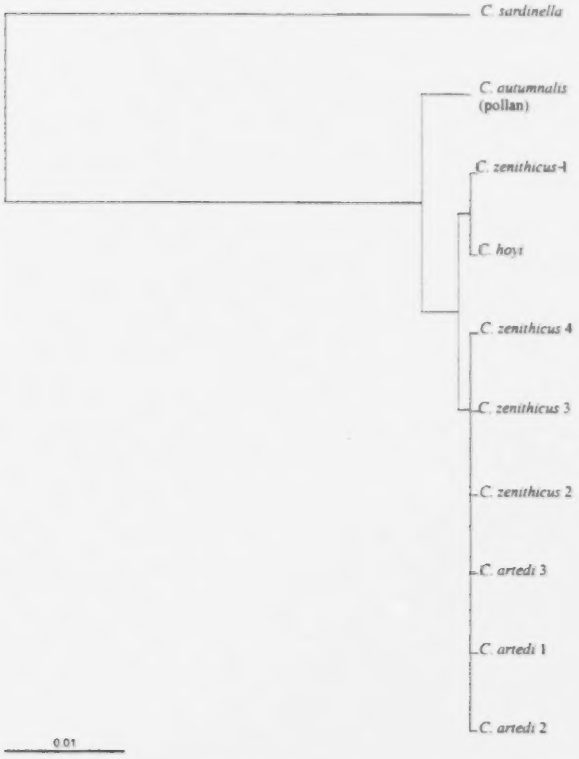



Figure 5. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND2 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter method (Kimura, 1980). Note that *C. autumnalis* (baicalensis) was omitted due to an incomplete sequence.

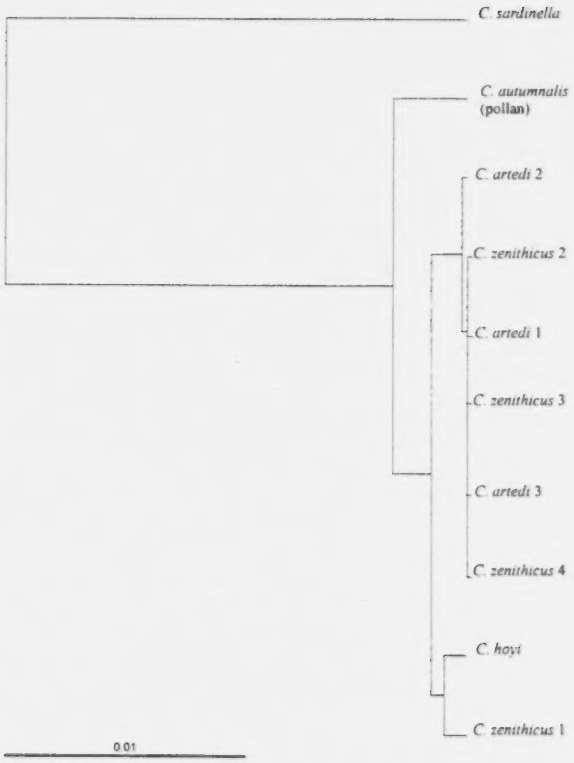



Figure 6. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COI mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequnece divergences estimated using the two-parameter method (Kimura, 1980). Note that *C. autumnalis* (baicalensis) was omitted due to an incomplete sequence.

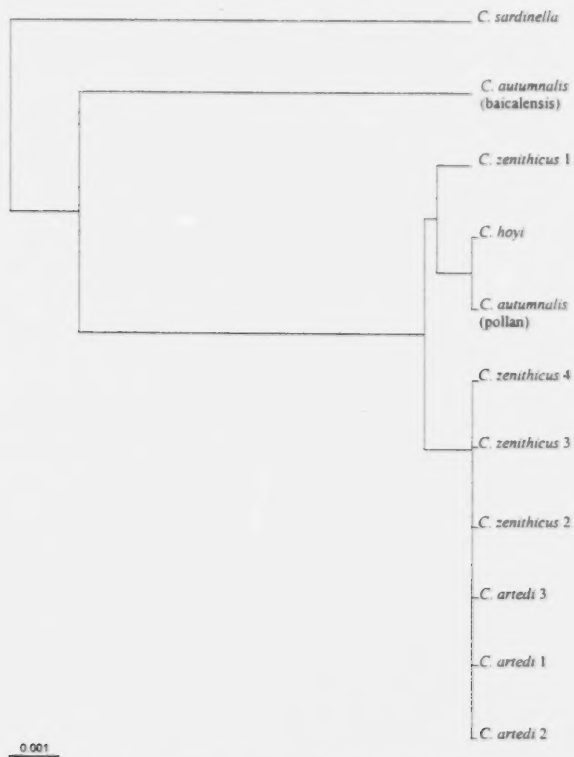



Figure 7. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COII

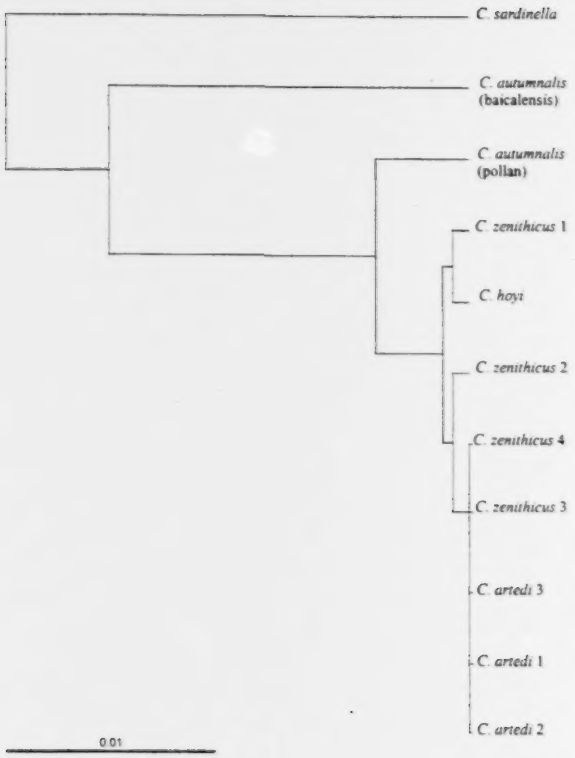



Figure 8. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ATPase6 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

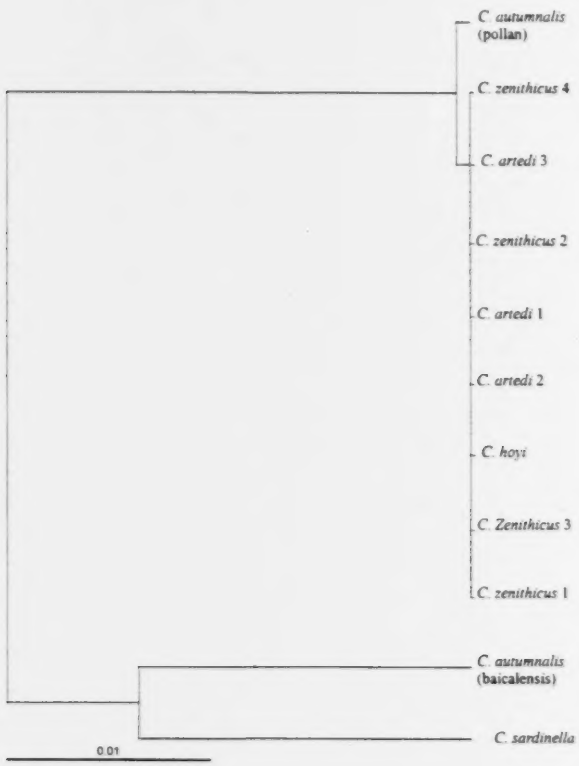



Figure 9. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the COIII mtDNAgene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence

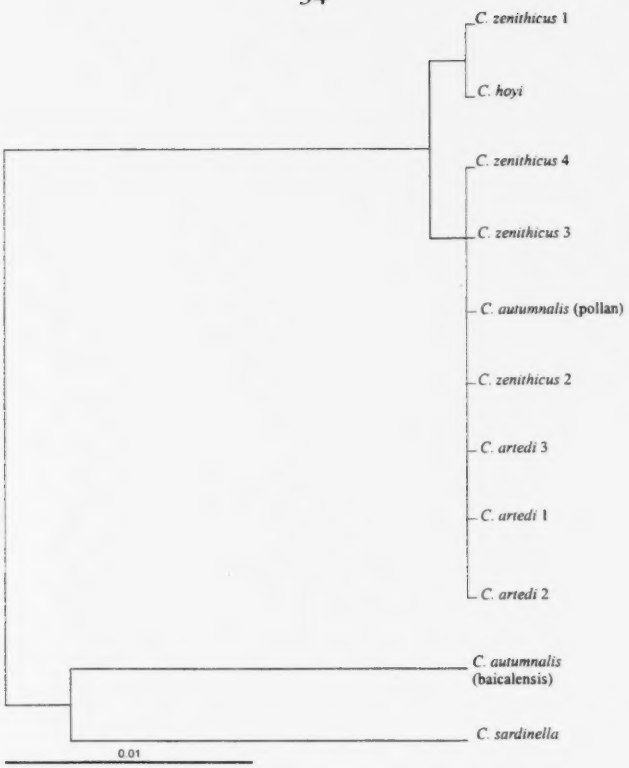



Figure 10. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND3 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

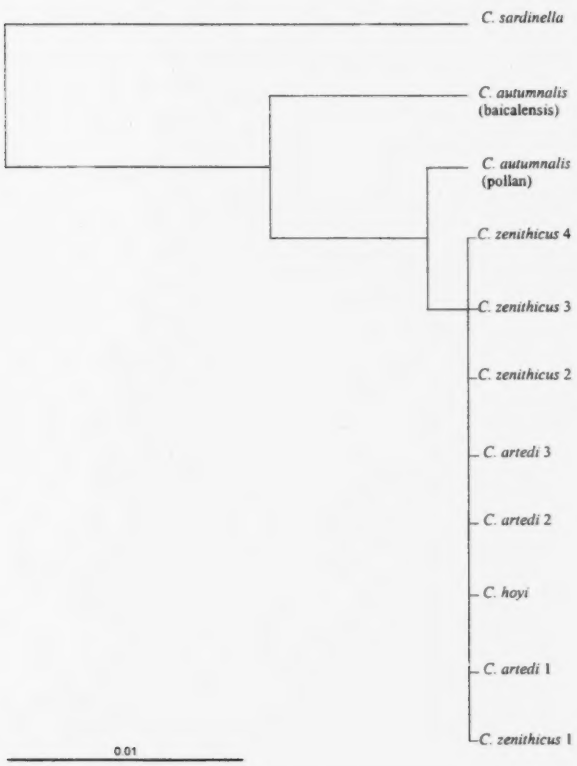



Figure 11. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND4L mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

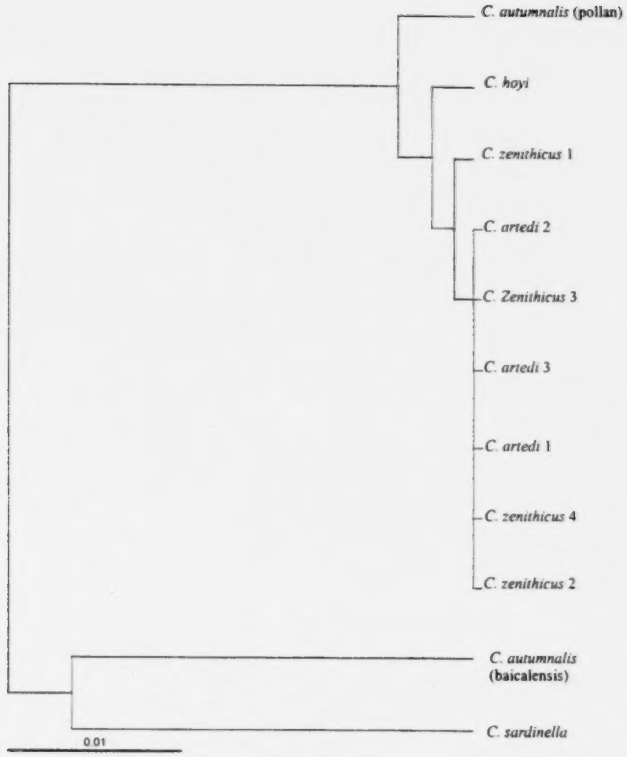



Figure 12. UPGMA (unweighted pair group with arithmetic mean) phylogram summarizing the ND4 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980)

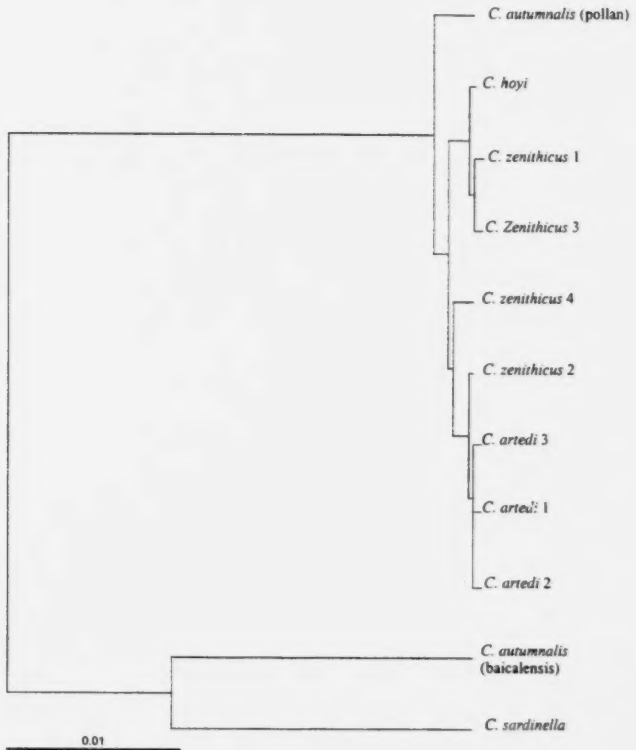



Figure 13. UPGMA (unweighted pair group with arithmetic mean) phylogram summarizing the ND5 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

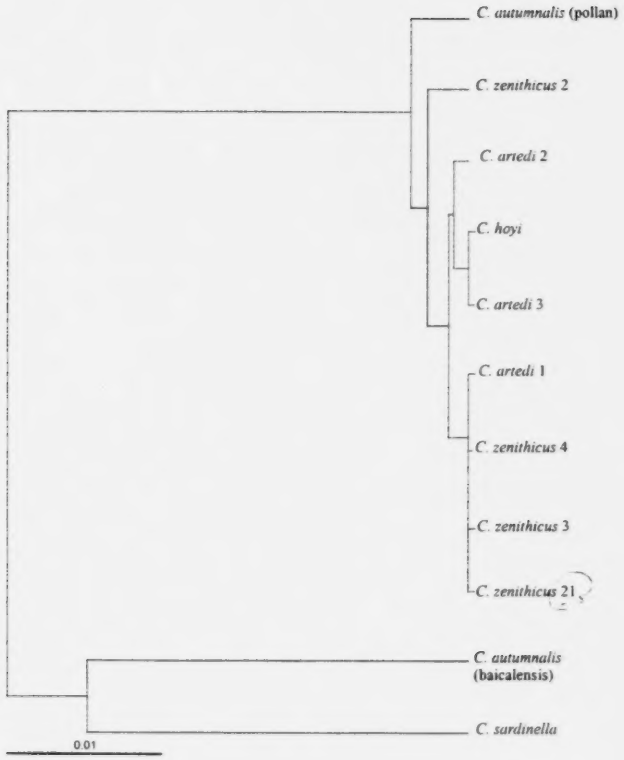



Figure 14. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the ND6 mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

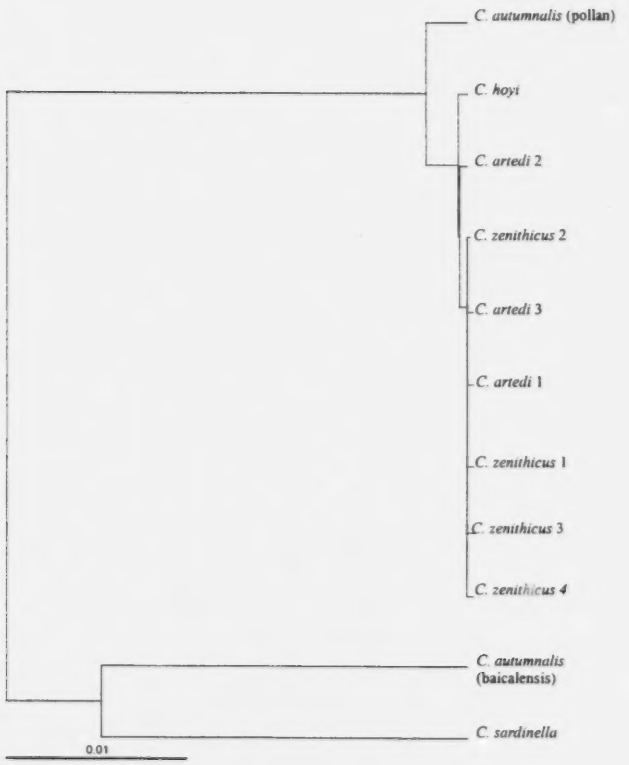



Figure 15. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the cyt *b* mtDNA gene relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence

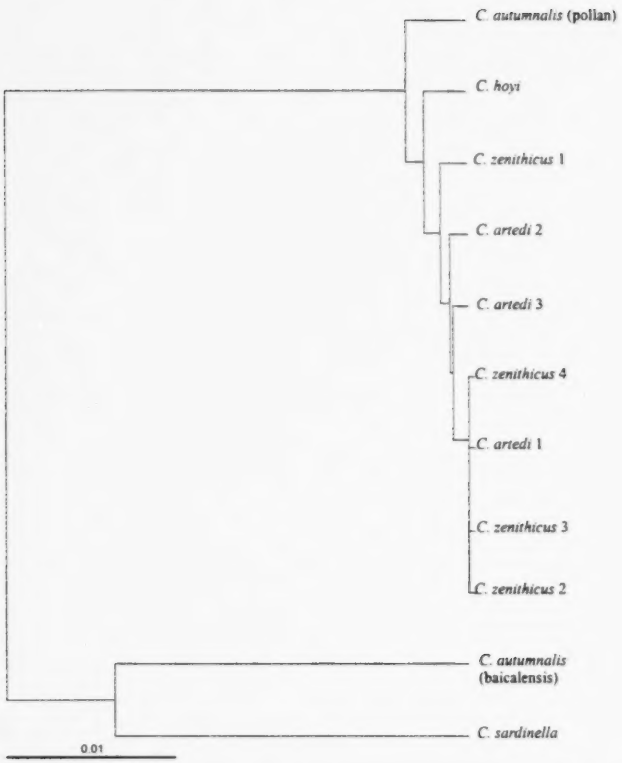



Figure 16. UPGMA (unweighted pair group method with arithmetic mean) phylogram summarizing the D-loop mtDNA relationships among the ciscoes studied. The phylogram was derived from a matrix of sequence divergences estimated using the two-parameter model (Kimura, 1980).

Appendix 1. Table detailing fish species and catch locations used in this study.

| Scientific name[1] | Common name | ID number [2] | Location |
|---|---|---|---|
| *Coregonus artedi* (1) | Lake cisco | 45763 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus artedi* (2) | Lake cisco | 45909 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus artedi* (3) | Lake cisco | 45941 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus zenithicus* (1) | Shortjaw cisco | 45943 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus zenithicus* (2) | Shortjaw cisco | 45923 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus zenithicus* (3) | Shortjaw cisco | 46023 | Lake Athapapaskow, Manitoba, Canada |
| *Coregonus zenithicus* (4) | Shortjaw cisco | 46272 | George Lake, Manitoba, Canada |
| *Coregonus hoyi* | Bloater | 33581 | Lake Michigan, United States |
| *Coregonus sardinella* | Least cisco | 43296 | North Killeak Lake, Alaska, United States |
| *Coregonus autumnalis* (pollan) | Irish pollan | 36303 | Lough Neagh, Ireland |
| *Cosregonus autumnalis migratorius* | Baikal cisco | 44418 | Lake Baikal, Russia |

1. Number in parenthesis indicates number accorded to the fish in the phylograms (e.g. *Coregonus artedi* (1) is *C. artedi* 1 in the phylograms).
2. Fish identification (ID) number and location code cross-reference the sample to the (FWI AFEAR) internal fish sample database.

Appendix 2. PCR and sequencing primers used in the analysis of coregonine mitochondrial genomes.

| + primers[1,2] | Sequence (5'→ 3')[3] | - primers[1,2] | Sequence (5'→ 3')[3] |
|---|---|---|---|
| 1. F16S | ACACCGAGAAGACATCCGTG | 20. 12S rRNA | AGGTTGAACGGGGGTTATCG |
| 2. F16S int | ACACAACCTTTACAGGCGGC | 21. R12S | TCCCTTTTCTGTCGCCCATAC |
| 3. ND1A | GCCTCGCCTGTTTACCAAAAACAT | 22. R16S | AACCTCTGAGAAAAAGGGC |
| 4. FNADH1 | AAGGTGTGCCTGAGATAACG | 23. ND1B | GGTATGGGCCCGAAAGCTTA |
| 5. FNADH2 | AAAAGCCGCCCTACTTACCG | 24. RND2 | AGGGTGCCAATGTCTTTG |
| 6. FNADH2 int | CAAAACTCTTGGTGCTTCC | 25. RNADH2 | ATCATAACGAAGGCGTGGGC |
| 7. FND2 | TGTTACACTGGGGCTTCTATC | 26. RCoxI | GCTAAGTGCTCAAGACACGC |
| 8. FCoI | ATGGGGTTACAATCCACCGC | 27. RCoI | GTGAGGGATGTGCCATTGTAGC |
| 9. FCoI int | TTTCCTCTATCTTAGGAGC | 28. Rcox2lav | ACGAGAATACCAAGATGG |
| 10. Fcox2cor | CAAGCCAACCGCATAACC | 29. Rcox3 int | GTAACAGTAACACCAGATGC |
| 11. FCoxII | ATAGTAGTCCCCGTGGAATCTC | 30. Rcox3lav | CAGAAAGAAACAGTAGCCAG |
| 12. ATPase6 | ATTTCTTTCCCAACCCCCTC | 31. RND4 | ATAGTGCTGAAGAAGCAAG |
| 13. Fcox3cor | CCTTTACCTACAAGAAAACG | 32. ND3/4B | TTTTGGTCCCTAAGACCAATGGAT |
| 14. ND3/4A | TTAATACGTATAAGTGACTTCCAA | 33. RND6 int | TAACGGGGGTTTTAGAGGG |
| 15. FNADH4 | TCTCTCAGCCCTTCTTTGC | 34. RND6 | GGAGTTCTCTGTTGTGCG |
| 16. FNADH5/6 | CTTATCCACCGAGAGAAGTCTGTTG | 35. RNADH5/6 | TGCGTCATTAGCAATCTTTAGGAG |
| 17. FND5 int | TCAACTGTTTATCGGCTGAG | 36. CytB2 | GAGCTACTAGGGCAGGCTCA |
| 18. CytB | GAAAAACCAYCGTTGTWATTCAACT | 37. H2A | CTGATAAATCGTTGGTCGG |
| 19. Char2 | CAAAACTCCAACTAACACGG | 38. H2B | GCCAGGAATAATTCACGAG |
| | | 39. DlpPhe | TATGCTTTAGTTAAGGCTACG |
| | | 40. HN20 | GTGTTATGCTTTAGTTAAGC |

1. Primers are designated by their 3' ends, corresponding to their relative positions along the *Coregonus lavaretus* mitochondrial genome (Miya and Nishida, 2000).
2. For relative positions in the mitochondrial genome, see Figure 1.
3. Positions with mixed bases are labeled with their IUB codes: W=A or T.